



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Studio degli Algoritmi di Estrazione del Segnale dei Rivelatori

Infrarosso dello Strumento NISP della Missione Spaziale Euclid

Relatore

Dr. Stefano Dusini

Correlatore

Dr. Luca Stanco

Laureanda

Elena Manao

Anno Accademico 2019/2020

Abstract

L'osservazione delle supenovae e lo studio della radiazione cosmica di fondo avvenuti negli ultimi anni hanno evidenziato un'accelerazione nell'espansione dell'universo, che può essere spiegata ipotizzando l'esistenza di una nuova forma di energia, l'energia oscura, o che la Teoria della Gravità cessi di valere su scale cosmologiche. La missione Euclid della European Space Agency (ESA) ha l'obiettivo di fare luce su questo fenomeno investigando l'evoluzione delle strutture cosmiche negli ultimi 10 miliardi di anni. Il satellite è composto da un telescopio a largo campo e due strumenti: il Visual Instrument (VIS) e il Near-Infrared Spectro-Photometer (NISIP). In questo lavoro di tesi saranno introdotti i nuovi algoritmi di estrazione del segnale per i rivelatori dello strumento NISP, verranno esposti i risultati ottenuti applicando gli algoritmi ai dati disponibili dalle campagne di test dello strumento e confrontati con i risultati attesi dalle simulazioni numeriche.

Indice

1	Introduzione	1
1.1	La missione Euclid	1
2	Lo strumento NISP	2
2.1	Detectors H2RG	2
2.2	Modalità di Lettura del Segnale	3
3	Algoritmo di Estrazione del Segnale	5
3.1	Stima del segnale accumulato tra due gruppi	5
3.1.1	Spazio delle Differenze tra Gruppi	5
3.1.2	Estimatore del Segnale	6
3.2	Verifiche	7
3.2.1	Accuratezza dell'Estimatore	7
3.2.2	Varianza del Flusso	7
3.2.3	Fattore di Qualità	7
4	Analisi e Risultati	9
4.1	Lettura delle Immagini e Analisi Preliminari	9
4.1.1	Il Formato FITS	9
4.1.2	Correzione e Selezione dei Dati	9
4.2	Stime di \hat{g}	11
4.3	Errori	12
4.4	Fattori di Qualità	13
5	Conclusioni	14
	Bibliografia	15

Capitolo 1

Introduzione

1.1 La missione Euclid

Euclid [1] è una missione della European Space Agency (ESA) progettata per compiere un'indagine estensiva del cielo per un periodo di sei anni e fare luce tra le teorie riguardanti l'espansione accelerata dell'Universo.

Tra le varie ipotesi avanzate, la più plausibile è che tale fenomeno sia da attribuire ad una sorgente completamente sconosciuta, che prende il nome di *energia oscura*.

La missione Euclid, nella sua indagine, misurerà la distribuzione di massa nell'universo in funzione del redshift sfruttando principalmente il *galaxy clustering* e il *weak lensing gravitazionale*.

Nel satellite si possono individuare principalmente due parti: il modulo di servizio (SVM) e il carico utile (PLM) [2].

Il modulo di servizio è costituito da tutti i componenti convenzionali di una navicella spaziale. I due componenti di maggior importanza nel SVM sono il sistema di puntamento e il sistema di regolazione termica, costruiti in modo da garantire un'alta precisione nelle attività di osservazione.

Il carico utile, rappresentato schematicamente in Figura 1.1, è costituito da un telescopio Korsch a tre specchi, che funge da interfaccia termica e meccanica con i due strumenti presenti a bordo: il Visible Imager (VIS) [3] e il Near-Infrared Spectrophotometer (NISP) [4]. Il VIS raccoglie immagini di galassie nella banda tra i 550 e i 900 nm ed è costituito da più componenti: un piano focale, un meccanismo di otturazione che chiude il cammino ottico e un'unità di calibrazione. Il NISP, invece, raccoglie immagini relative a segnali di lunghezza d'onda appartenente alla regione del vicino infrarosso, è progettato per lavorare in due modalità, spettroscopia slitless e imaging fotometrico. Lo strumento è in grado di passare da una modalità all'altra per mezzo del sistema ottico implementato.

Gli strumenti sono progettati per eseguire una ricognizione che, nell'intera durata della missione, coprirà circa il 36% del cielo.

Il lavoro di tesi è strutturato in questo modo: nel capitolo 2 sono descritte le principali caratteristiche dello strumento NISP, ponendo l'attenzione principalmente sulle modalità di acquisizione dati, nel capitolo 3 è introdotto il nuovo estimatore per il flusso e nel capitolo 4 sono illustrati i risultati ottenuti.

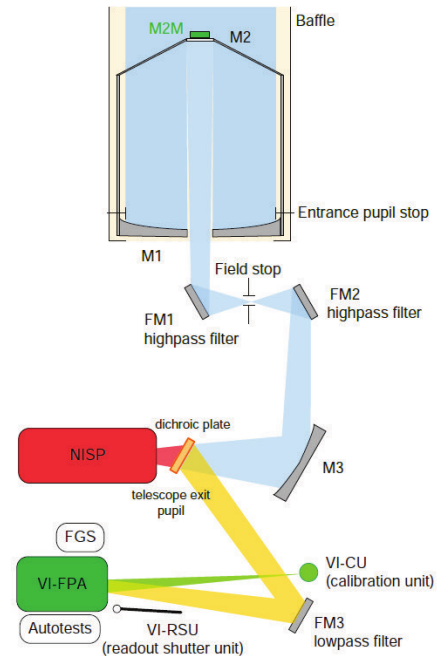


Figura 1.1: Schema Payload Module.

Capitolo 2

Lo strumento NISP

Il *Near-Infrared Spectro-Photometer* (NISP) è uno spettro-fotometro che esegue spettroscopia slitless e imaging fotometrico nella banda del vicino infrarosso. I principali sottosistemi dello strumento sono:

- il sistema di filtri e grism necessari alle osservazioni in modalità fotometrica e spettroscopica;
- il piano focale composto da 16 detectors HAWAII-2RG (H2RG);
- l'unità elettronica;
- l'unità di calibrazione;

Di seguito vengono illustrate le principali caratteristiche dei detectors H2RG e la modalità di lettura del segnale.

2.1 Detectors H2RG

Il piano focale dello strumento NISP è composto da una matrice 4x4 di detector HAWAII-2RG.¹

I detector H2RG sono costituiti da cristalli di HgCdTe, un materiale semiconduttore in grado di assorbire fotoni per effetto fotoelettrico in un range di lunghezze d'onda comprese tra i 1,24 e 18 μm . I fotoni assorbiti e convertiti in elettroni possono essere poi raccolti grazie alla presenza di un campo elettrico all'interno del cristallo, ottenuto applicando una differenza di potenziale ai suoi capi. Per fare sì che la corrente prodotta rientri nei range richiesti, i cristalli dei detector H2RG sono disposti a formare una giunzione PN, in modo che il detector così costruito sia un fotodiodo [5]. Per consentire la raccolta della carica elettrica generata per effetto fotoelettrico si opera in regime di polarizzazione inversa. In particolare si applica una differenza di potenziale nota ai capi della giunzione e si misura la variazione di quest'ultima nel tempo in uno dei due capi. Una volta che il segnale viene completamente raccolto, lo si digitalizza tramite un ADC².

Per ottenere immagini bidimensionali, è necessario associare un gran numero di fotodiodi in modo da formare una matrice 2D di pixel. Ciò è solitamente fatto impiegando dei sensori di tipo CMOS. Questi sensori hanno forma di una matrice 2D di pixel di piccole dimensioni; ogni pixel contiene un fotodiodo e un amplificatore che converte la carica in un segnale in tensione. Ciò consente di leggere ogni pixel in modo indipendente degli altri.

I detector H2RG sono costituiti da 2048x2048 pixel con un'elettronica di lettura (ADC) con codifica del segnale a 16 bit. I pixel fotosensibili sono solamente 2040x2040 e sono circondati da 4 righe e 4 colonne di reference pixel, che non sono collegati ai fotodiodi ma contengono una semplice capacità che mima i fotodiodi dei pixel operativi. I reference pixel sono di fondamentale importanza per tenere traccia di eventuali bias e variazioni di temperatura che possono emergere nelle esposizioni più lunghe.

¹HAWAII è l'acronimo di *HgCdTe Astronomical Wide Area Infrared Imager*, il 2 sta ad indicare che il detector restituisce un array bidimensionale, R sta per *reference pixels* e G invece per *guide window capability*

²Analog to Digital Converter

2.2 Modalità di Lettura del Segnale

Il segnale accumulato da ogni pixel durante un'esposizione viene misurato ad intervalli di tempo costanti. Ogni lettura del segnale è non-distruttiva, ciò significa che, durante un'acquisizione, il detector rileva il segnale accumulato dal pixel nel tempo a partire dall'istante in cui ha inizio l'esposizione; in questo caso parliamo di segnale integrato. Al termine dell'esposizione il detector viene resettato e ha inizio una nuova acquisizione.

Esistono diverse strategie compatibili con l'elettronica dei detector per trattare il segnale integrato. La prima consiste nel trasferire solo i segnali relativi alle letture che avvengono immediatamente dopo il reset e alla fine dell'esposizione. In questo modo tutte le letture intermedie non vengono considerate. Il segnale integrato rilevato dal singolo pixel risulta essere, indicando con $S(t)$ la misura del segnale al tempo t :

$$\Delta S = S(t) - S(t_0) \quad (2.1)$$

L'incertezza su ΔS è legata sia all'errore di lettura (readout noise) che all'errore poissoniano sul conteggio e risulta essere

$$\sigma = \sqrt{2\sigma_R^2 + N} \quad (2.2)$$

L'errore di lettura è pari a $\sigma_R = 6,2$ ADU [6] e, per le illuminazioni attese, N va da $100 e^-$ a $1000 e^-$.³

Questo metodo, al crescere di N , porta ad avere errori di oltre il 10%; perciò si preferisce usare altri approcci, che presentano errori minori.

In particolare, la modalità di acquisizione dati dei rivelatori del NISP combina due strategie: l'acquisizione *Flower-M* e quella *Up the Ramp* (UTR).

L'acquisizione in modalità *Flower-M* riduce il contributo dato dall'errore di lettura dei pixel; invece di acquisire solo due immagini, come nel caso precedente, vengono acquisiti un gruppo di M immagini all'inizio e un altro alla fine dell'esposizione. Il segnale di ogni gruppo, G [7], è dato dalla media dei segnali letti nei singoli frame. Quindi, chiamando n_f il numero di frame per gruppo:

$$G = \frac{1}{n_f} \sum_{i=1}^{n_f} S_i \quad (2.3)$$

e l'errore sulla misura, come previsto, è ridotto, rispetto al metodo precedente, di un fattore \sqrt{M}

$$\sigma_{Flower_M} = \frac{\sigma}{\sqrt{M}} \quad (2.4)$$

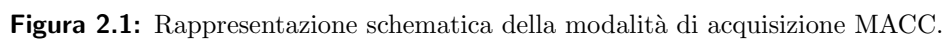
L'acquisizione in modalità UTR consiste, invece, nell'acquisire ogni frame letto dall'inizio alla fine dell'esposizione. L'errore tra due frame consecutivi è sempre la somma quadratica dell'errore di lettura e dell'incertezza poissoniana sul conteggio, quindi due letture consecutive sono correlate a causa del contributo di errore dovuto al numero di fotoni incidenti. Usando questa modalità, il flusso può essere stimato con un fit lineare delle letture acquisite. Lo svantaggio di questo metodo è legato alla considerevole mole di dati raccolti in una singola esposizione, aspetto da tenere in considerazione soprattutto in relazione alle capacità dello strumento, che ha un limite telemetrico di 250 Gbits al giorno, che impone che solo una quantità limitata di dati possa essere trasferita a terra.

La modalità di acquisizione del segnale che combina i vantaggi del metodo *Flower-M* e dell'acquisizione UTR prende il nome di *Multiple Accumulated Sampling* (MACC).

La MACC consiste nell'acquisizione di un numero n_g di gruppi ciascuno di n_f immagini, separate da n_d immagini non trasferite (*dropped*). In Figura 2.1 ne riportiamo una rappresentazione grafica.

L'acquisizione in modalità MACC consente sia di ridurre il contributo di errore dato dall'errore di lettura secondo la (2.4), sia di eseguire un fit lineare dei segnali dei singoli gruppi, ottenuti secondo la (2.3) per ricavare una stima del flusso. La notazione usata per specificare le caratteristiche dell'acquisizione è $MACC(n_g, n_f, n_d)$.

³[e^-]=Elettroni



I parametri della MACC sono scelti in modo da ottimizzare la presa dati in relazione alla durata dell'esposizione e ai tempi necessari al processamento dei dati. In particolare, il tempo che intercorre tra una lettura e la successiva, $t_f = 1,45408$ s [5].

Capitolo 3

Algoritmo di Estrazione del Segnale

In questo capitolo vengono spiegati i passaggi svolti per la definizione di un nuovo estimatore del segnale raccolto tra due gruppi consecutivi, \hat{g} [8], e illustrate alcune verifiche della sua validità eseguite tramite simulazioni Monte Carlo.

Le formule presentate in questo capitolo e tutte quelle che verranno esposte in seguito possono essere applicate direttamente ai dati grezzi e sono perciò espresse in ADU.

3.1 Stima del segnale accumulato tra due gruppi

3.1.1 Spazio delle Differenze tra Gruppi

A causa della lettura non-distruttiva dei pixel, i dati raccolti durante un'esposizione sono intrinsecamente correlati, perciò la matrice di covarianza tra i dati relativi a due gruppi diversi risulta avere tutte le entrate diverse da zero. Ciò comporta la necessità di implementare metodi numerici che permettano di invertire matrici di questo tipo, il che eccede le tipiche risorse di calcolo presenti su un satellite come Euclid.

Un modo per rimuovere gran parte delle correlazioni e semplificare la forma della matrice di covarianza consiste nel lavorare nello spazio delle differenze tra gruppi, che indichiamo con $\Delta G_k = G_{k+1} - G_k$. In questo caso, per una $\text{MACC}(n_g, n_f, n_d)$, la maggior parte delle entrate non diagonali della matrice di covarianza sono nulle, a parte per i termini di indici $(k, k \pm 1)$, con k che varia tra 1 e $n_g - 1$, che risultano tuttavia essere trascurabili a partire da flussi superiori ai 5 e⁻/s. I termini diagonali della matrice invece hanno la seguente forma:

$$D_{kk}(\Delta G_k, g) = \frac{2\sigma_R^2}{nf} + \frac{(n_d + 1)}{n_f + n_d} \frac{g}{f_e} + \frac{(n_f - 1)(2n_f - 1)}{3n_f(n_f + n_d)} \frac{g}{f_e} \quad (3.1)$$

dove con g indichiamo il segnale accumulato tra due gruppi successivi e con f_e il guadagno dello strumento, che nel nostro caso è pari a $f_e = \frac{1}{1.95}$ e⁻/ADU [6].

Il primo termine dell'espressione corrisponde all'errore di lettura, che consideriamo essere non correlato da frame a frame, il secondo all'errore sul conteggio tra l'ultimo frame del gruppo k e il primo del gruppo $k + 1$ e il terzo termine è dovuto alle correlazioni che emergono dalla MACC. É possibile riscrivere l'equazione (3.1) in modo più semplice, raccogliendo all'interno di un coefficiente, α , le correlazioni di Poisson. L'espressione dell'errore diventa dunque:

$$D_{kk}(\Delta G_k, g) = \frac{2\sigma_R^2}{nf} + (1 + \alpha) \frac{g}{f_e} \quad \text{con} \quad \alpha = \frac{1 - n_f^2}{3n_f(n_f + n_d)} \quad (3.2)$$

Notiamo che α dipende unicamente dalle caratteristiche della MACC, ed è dunque comune ad ogni pixel.

3.1.2 Estimatore del Segnale

Uno dei metodi più utilizzati per stimare parametri a partire dai dati è la massima verosimiglianza; in questo caso è necessario adattare tale metodo ad un modello che descrive la lettura di immagini in maniera non-distruttiva e che abbia una soluzione analitica, in modo da poterla implementare facilmente nell'elettronica di bordo.

Nell'ipotesi che il detector abbia una risposta perfettamente lineare, sia termicamente stabile e non sia soggetto ad alcun effetto che possa variare durante l'esposizione la funzione verosimiglianza è la seguente:

$$\mathcal{L} = \prod_{i=1}^{n_g-1} \frac{1}{\sqrt{2\pi\sigma_{eff}^2(\Delta G_i, g)}} \exp\left(-\frac{(\Delta G_i - g)^2}{2\sigma_{eff}^2(\Delta G_i, g)}\right) \quad (3.3)$$

dove l'errore al quadrato su ΔG_i è definito come:

$$\sigma_{eff}^2(\Delta G_i, g) = D_{kk}(\Delta G_i, g) \quad (3.4)$$

in modo da tenere conto sia dell'errore di lettura che dell'errore sul conteggio, che delle correlazioni di Poisson.

L'estimatore del segnale, \hat{g} , corrisponde al valore di g che massimizza \mathcal{L} . Si procede quindi calcolando il logaritmo di \mathcal{L} e cercandone il massimo rispetto a g . In particolare si dovrà calcolare la derivata prima di $l = -2\log\mathcal{L}$ rispetto a g e trovare il valore di g per cui si annulla.

Ripercorriamo i passaggi; l ha forma:

$$l = -2 \log\mathcal{L} = \sum_{i=0}^{n_g-1} \frac{(\Delta G_i - g)^2}{(1 + \alpha)\frac{g}{f_e} + \frac{2\sigma_R^2}{nf}} + (n_g - 1)\log\left[(1 + \alpha)\frac{g}{f_e} + \frac{2\sigma_R^2}{nf}\right] + (n_g - 1)\log 2\pi \quad (3.5)$$

dalla derivata prima di l si ricava l'espressione dell'estimatore, \hat{g} , che risulta essere

$$\hat{g} = \frac{1 + \alpha}{2f_e} \left[\sqrt{1 + \frac{4f_e^2 \sum_{i=1}^{n_g-1} (\Delta G_i + \beta)^2}{(n_g - 1)(1 + \alpha)^2}} - 1 \right] - \beta \quad (3.6)$$

mentre dalla derivata seconda si ricava la stima di $\sigma_{\hat{g}}$.

$$\sigma_{\hat{g}}^2 = \frac{2(\hat{g} + \beta)^2}{(n_g - 1) \left(1 + \frac{2f_e}{1 + \alpha}(\hat{g} + \beta)\right)} \quad (3.7)$$

dove è stato introdotto il coefficiente β per semplificare la notazione. β risulta essere

$$\beta = \frac{2\sigma_R^2 f_e}{n_f(1 + \alpha)} \quad (3.8)$$

È importante notare come la funzione verosimiglianza usata costituisca un'approssimazione, in quanto i ΔG hanno, in realtà, distribuzione poissoniana.

Per mettere in evidenza l'efficacia del nuovo estimatore appena introdotto si può andare a vedere l'effetto che si ha nel caso in cui vengano trascurate le correlazioni di Poisson, quindi ponendo $\alpha = 0$ in (3.6) e (3.7), o nel caso in cui il fattore di normalizzazione della funzione di verosimiglianza non dipenda da g , il che è ovviamente falso dato che abbiamo a che fare con variabili casuali con distribuzione di Poisson. La espressione di \hat{g} in questo caso è:

$$\hat{g}_{\text{no likelihood}} = \sqrt{\frac{\sum_{i=0}^{n_g-1} (\Delta G_i - \beta)^2}{n_g - 1}} - \beta \quad (3.9)$$

mentre $\sigma_{\hat{g}}$ è esattamente $\sigma_{eff}(\Delta G_i, g)$.

3.2 Verifiche

3.2.1 Accuratezza dell'Estimatore

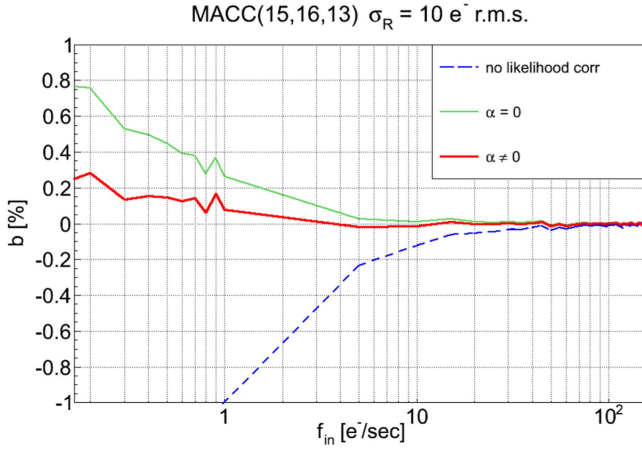


Figura 3.1: Bias dell'estimatore del flusso (3.10) in funzione del flusso simulato per le tre diverse stime di \hat{g} prese in considerazione. In rosso è rappresentato il bias rispetto alla stima di \hat{g} calcolata con la (3.6), in verde è rappresentato il caso $\alpha = 0$ e in blu tratteggiato il caso denominato "no likelihood" (3.9) [8].

3.2.2 Varianza del Flusso

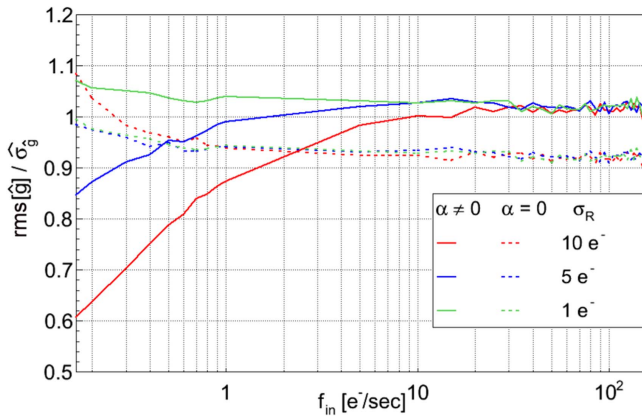


Figura 3.2: Rapporti $\text{rms}[\hat{g}]/\sigma_{\hat{g}}$ nei casi $\alpha \neq 0$ (linee continue) e $\alpha = 0$ (linee tratteggiate) per tre diversi valori di σ_R (readout noise) [8].

importanti. Inoltre, quando le correlazioni di Poisson sono trascurate l'errore risulta sistematicamente sovrastimato del 10% indipendentemente dal valore del flusso.

3.2.3 Fattore di Qualità

Il metodo utilizzato per ottenere \hat{g} ha il pregio di fornire un altro estimatore statistico, il fattore di qualità, che ha lo scopo di verificare se i dati seguano la distribuzione poissoniana convoluta con una gaussiana. Il fattore di qualità, QF è definito come:

$$\text{QF} = \sum_{i=1}^{n_g-1} \frac{(\Delta G_i - g)^2}{n_g - 1} \Big|_{g=\hat{g}_x} \quad (3.11)$$

¹Con la larghezza della distribuzione di \hat{g} intendiamo l' $\text{rms}[\hat{g}]$

L'accuratezza della stima del flusso ottenuta dall'estimatore \hat{g} può essere verificata computando il *bias*, ossia la differenza relativa dal flusso simulato:

$$b = \frac{f_{in} - \hat{g}/t_g}{f_{in}} \quad (3.10)$$

dove f_{in} è il flusso simulato e $t_g = (n_f + n_d)t_f$ il tempo che intercorre tra l'inizio di due gruppi consecutivi. Nel grafico in Figura 3.1 sono riportati i risultati di una simulazione di 10 000 esposizioni non-distruttive in un range di flussi che va da 0,1 e⁻/s a 150 e⁻/s.

Notiamo che il bias di \hat{g} , calcolato secondo la (3.6), è inferiore allo 0,3% per tutto il range di flussi considerato. Nel caso in cui si trascurano le correlazioni di Poisson ($\alpha = 0$) il bias arriva fino allo 0,8% e, nel caso in cui, invece, si considera $\hat{g}_{\text{no likelihood}}$ (3.9), il bias assume valori ben superiori all'1% per flussi molto bassi.

Per verificare l'accuratezza della stima di $\sigma_{\hat{g}}$ (3.7) è possibile studiare come varia il rapporto tra la larghezza di della distribuzione di \hat{g}^1 e la stima ottenuta dal metodo della massima verosimiglianza. In particolare, nel grafico sono rappresentati i rapporti $\text{rms}[\hat{g}]/\sigma_{\hat{g}}$ in funzione dei flussi simulati per tre diversi valori di σ_R sia nel caso $\alpha \neq 0$ che per $\alpha = 0$.

Se le ipotesi fatte per il computo dell'errore sono corrette, allora l' rms osservato è uguale al risultato analitico, e il rapporto è quindi pari a uno. Notiamo che il modello è in ottimo accordo con i dati osservati per flussi superiori a 10 e⁻/s mentre a flussi più bassi c'è un'evidente sovrastima di $\sigma_{\hat{g}}$. Tale sovrastima è causata dalle correlazioni tra i frame di due gruppi consecutivi, che abbiamo trascurato sebbene a flussi bassi diventano

dove \hat{g}_x è il valore di g nel minimo del QF, che ha forma

$$\hat{g}_x = \sqrt{\sum_{i=1}^{n_g-1} \frac{(\Delta G_i + \beta)}{n_g - 1}} - \beta \quad (3.12)$$

Il QF (3.11), dato che è un prodotto di densità di probabilità gaussiane, si distribuisce come il χ^2_{th} a n gradi di libertà. In questo caso il numero di gradi di libertà è $n = n_g - 2$ [8], dove n_g è il parametro della MACC che indica il numero di gruppi rilevati durante l'esposizione.

Ancora una volta, in Figura 3.3, vediamo come la simulazione eseguita utilizzando le equazioni (3.6) e (3.7) (linea blu) segue strettamente la distribuzione attesa (zona ombreggiata), confermando che gli errori sono stimati correttamente. Ciò non accade quando si trascurano le correlazioni poissoniane e si pone, quindi, $\alpha = 0$ (linea verde).

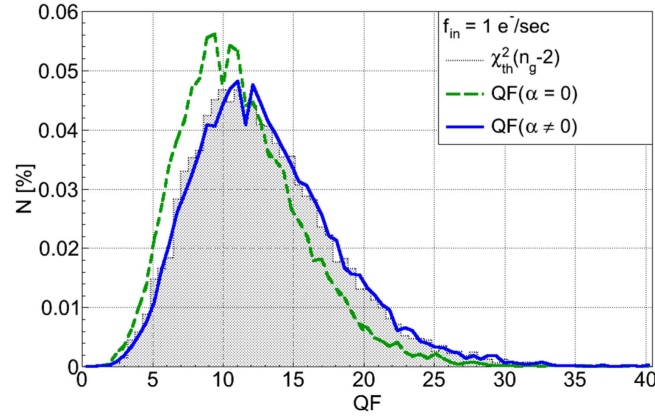


Figura 3.3: Distribuzione dei fattori di qualità ottenuta dalla simulazione eseguita considerando un flusso incidente di $1 \text{ e}^-/\text{s}$. In grigio ombreggiato è rappresentata la funzione $\chi^2_{th}(n_g - 2)$, in blu la distribuzione dei QF nel caso $\alpha \neq 0$ e in verde tratteggiato il caso $\alpha = 0$ [8].

Capitolo 4

Analisi e Risultati

In questo capitolo sono illustrate le principali caratteristiche dei file in formato FITS, il modo in cui è stata svolta l'analisi dati e i risultati ottenuti.

4.1 Lettura delle Immagini e Analisi Preliminari

4.1.1 Il Formato FITS

I dati acquisiti da NISP sono prima scritti sulla Mass Memory e successivamente convertiti in file FITS¹ [9]. Il formato FITS è ampiamente usato in astronomia per il trasporto, l'analisi e l'archivio di set di dati. I file a nostra disposizione hanno estensione ENGRAW e contengono:

- un numero n_g di immagini contenenti il segnale relativo ai singoli gruppi acquisiti, ossia la media del segnale acquisito dagli n_f frame, denotato G nell'equazione (2.3);
- l'immagine *science*, nella quale è rappresentato il segnale integrato dall'unità di bordo del satellite;
- l'immagine *chi2*, nella quale sono riportati i fattori di qualità.
- due tabelle binarie contenenti le telemetrie di servizio.

Tutte le immagini vengono lette come matrici di dimensione 2048x2048 e, da queste, vengono costruiti gli istogrammi che rappresentano la distribuzione dei dati acquisiti da NISP presenti nelle immagini.

In questo lavoro prendiamo in analisi tre file relativi a tre diverse illuminazioni del detector. Tutti e tre i file sono stati acquisiti in modalità MACC(10,16,11).

4.1.2 Correzione e Selezione dei Dati

Per questo lavoro di tesi abbiamo analizzato dei dati raccolti con un rivelatore *engineering grade* la cui risposta è risultata essere non completamente uniforme e soggetta a qualche difetto. Per questa ragione, prima di implementare gli algoritmi per la stima del segnale è necessario apportare una correzione ai dati grezzi e selezionare la zona del detector i cui pixel rispondono nel modo più uniforme possibile. Nella definizione dell'estimatore (sezione 3.1.2), infatti, abbiamo ipotizzato che il detector sia lineare e stabile; è quindi necessario individuare la zona del detector che meglio soddisfa queste caratteristiche.

La correzione che apportiamo alle immagini relative ai singoli gruppi è computata a partire dai dati rilevati dai reference pixel e ha lo scopo di omogeneizzare la risposta del detector.

Come spiegato in sezione 2.1, i pixel attivi sono circondati da quattro righe e quattro colonne di reference pixel. Il segnale raccolto dai reference pixel è meno intenso rispetto a quello dei pixel attivi, come si può notare in Figura 4.1, e viene impiegato per misurare il rumore indotto dall'elettronica di

¹Flexible Image Transport System

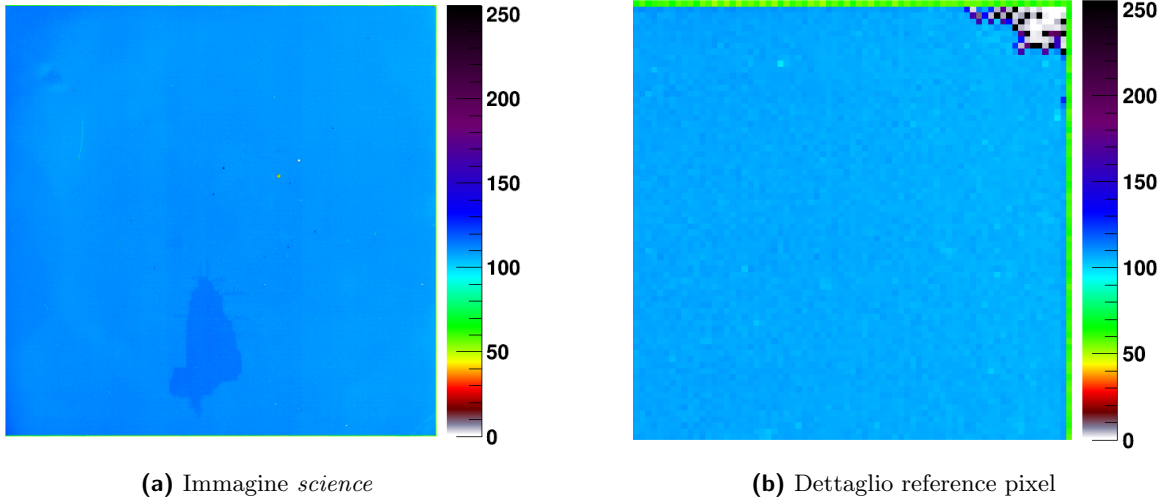


Figura 4.1: Figura (a): Immagine *science* di uno dei file analizzati. La zona più scura nella parte bassa dell'immagine corrisponde ai pixel che rispondono in modo diverso dal resto del detector. Figura (b): Dettaglio dell'angolo in alto a destra dell'immagine *science* di uno dei file analizzati.

lettura in modo da poterlo sottrarre, pixel per pixel, al segnale della parte sensibile del rivelatore. In particolare, per questa correzione vengono computati separatamente due contributi: l'uno relativo alle righe di reference pixel collocate nella parte alta e bassa del detector e l'altro relativo ai restanti reference pixel nelle regioni di destra e di sinistra. Questi contributi vengono mediati e sottratti ai pixel attivi [6].

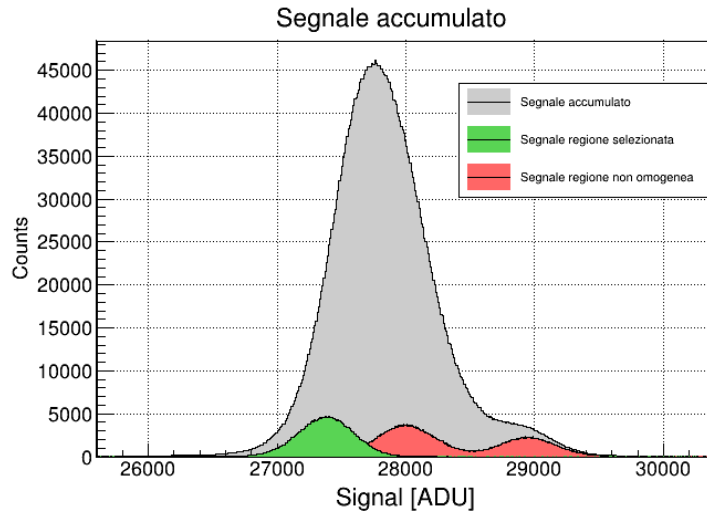


Figura 4.2: Segnale integrato durante l'esposizione. Nell'istogramma grigio è rappresentato il segnale di tutti i pixel attivi, quello verde si riferisce ai pixel sui quali viene svolta l'analisi e quello rosso è invece l'istogramma dei pixel relativi alla regione non omogenea al resto del detector.

Per quanto riguarda, invece, l'uniformità della risposta del detector, osserviamo l'immagine *science* di uno dei file che prendiamo in analisi (Figura 4.1). In questa immagine siamo in grado di individuare una regione in cui la risposta del detector è chiaramente diversa dal resto. Questo comportamento anomalo è chiaramente visibile in Figura 4.1. Inoltre, calcolando la distribuzione del segnale (Figura 4.2) notiamo un eccesso di segnale nella parte destra della distribuzione (istogramma grigio). Selezionando i pixel della zona del detector circostante la regione che abbiamo individuato avere un comportamento anomalo e disegnandone l'istogramma otteniamo l'istogramma rappresentato in rosso, che presenta un picco in corrispondenza dell'eccesso di segnale precedentemente individuato.

Il terzo istogramma, rappresentato in verde in Figura 4.2, è l'istogramma del segnale relativo alla

regione di pixel che abbiamo selezionato per l'analisi. La regione scelta è quella compresa tra le righe numero 1500 e 2000 e le colonne numero 1500 e 2000. La distribuzione del segnale di quest'ultimo istogramma è compatibile con una gaussiana e non presenta disomogeneità evidenti; ci concentriamo quindi su questa regione per testare l'efficacia degli algoritmi definiti in 3.1.2.

4.2 Stime di \hat{g}

Confrontiamo le stime di \hat{g} che otteniamo applicando (3.6) e (3.7) con la stima che possiamo ricavare dall'immagine *science*. Nell'immagine *science*, infatti, è rappresentata la stima di \hat{g} ottenuta dagli algoritmi di estrazione del segnale implementati nell'elettronica di bordo dello strumento moltiplicata per un fattore $n_g - 1$ e, di conseguenza, la stima di \hat{g} che otteniamo applicando questi algoritmi deve essere compatibile con il valor medio della distribuzione del segnale in *science* diviso per un fattore $n_g - 1$.

Le stime di \hat{g} e $\sigma_{\hat{g}}$ riportate in Tabella 4.1 sono state ottenute applicando gli algoritmi pixel per pixel e calcolando poi la media della distribuzione degli estimatori così computati. Come possiamo notare in Figura 4.3 i valori di \hat{g} e $\sigma_{\hat{g}}$ si distribuiscono secondo una gaussiana; tutte le stime riportate di seguito sono il valore di aspettazione della gaussiana che meglio interpola gli istogrammi. Oltre alle stime di \hat{g} , riportiamo le stime del flusso incidente ($f_{in} = \frac{\hat{g}}{t_g} f_e$) convertito in elettroni al secondo per tutti e tre i casi discussi in precedenza. Come previsto dalle simulazioni, avendo a che fare con flussi superiori ai $5 \text{ e}^-/\text{s}$, non notiamo differenze rilevanti nelle stime di \hat{g} , e quindi dei flussi, ma constatiamo invece una pesante sovrastima dell'errore nel caso in cui assumiamo che il fattore di normalizzazione della funzione verosimiglianza non dipenda da g . Le stime di \hat{g} e \hat{g}_0 , infatti, hanno un errore pari rispettivamente a circa il 0,7% e 0,8%, mentre \hat{g}_{nl}^2 ha un errore del 2,3%, circa il triplo degli altri due casi.

Illuminazione	\hat{g} [ADU]	f_{in} [e^-/s]	\hat{g}_0 [ADU]	f_{in_0} [e^-/s]	\hat{g}_{nl} [ADU]	$f_{in_{nl}}$ [e^-/s]
1	3044 ± 23	$41,00 \pm 0,31$	3044 ± 25	$41,01 \pm 0,33$	3044 ± 69	$41,01 \pm 0,93$
2	2388 ± 20	$32,16 \pm 0,27$	2387 ± 23	$32,16 \pm 0,31$	2388 ± 62	$32,17 \pm 0,82$
3	2569 ± 21	$34,60 \pm 0,28$	2569 ± 24	$34,60 \pm 0,32$	2570 ± 63	$34,61 \pm 0,85$

Tabella 4.1: Stime di \hat{g} in ADU e del flusso incidente convertito in e^-/s ottenute applicando gli algoritmi introdotti in 3.1.2 per tutti e tre i casi discussi. La stima con pedice *nl* si riferisce all'equazione (3.9). L'errore di f_{in} è calcolato per propagazione degli errori, considerando un errore sul tempo pari a 0,01 s.

In Tabella 4.2 riportiamo invece la stima di \hat{g} che ricaviamo dalla distribuzione del segnale contenuto nell'immagine *science* (Figura 4.4) e la compatibilità³ con la stima di \hat{g} ottenuta applicando l'algoritmo. In tutti e tre i casi \hat{g}_{sci} è in ottimo accordo con le stime ricavate dai dati grezzi.

Illuminazione	μ_{sci} [ADU]	\hat{g}_{sci} [ADU]	\hat{g} [ADU]	Compatibilità
1	27375 ± 217	3042 ± 24	3044 ± 23	0,07
2	21474 ± 182	2386 ± 20	2388 ± 20	0,05
3	23098 ± 199	2566 ± 22	2569 ± 21	0,08

Tabella 4.2: Per ogni illuminazione riportiamo la media del segnale accumulato dal detector durante tutta l'esposizione (μ_{sci}), la stima di \hat{g} che ricaviamo dall'istogramma del segnale integrato, $\hat{g}_{sci} = \frac{\mu_{sci}}{n_g - 1}$, \hat{g} e la compatibilità tra le due stime.

² \hat{g}_{nl} sta ad indicare $\hat{g}_{no \text{ likelihood}}$, definito in (3.9).

³La compatibilità tra due stime è definita come $\lambda = \frac{|a-b|}{\sqrt{\sigma_a^2 + \sigma_b^2}}$, con a e b le due stime delle quali si vuole ottenere la compatibilità.

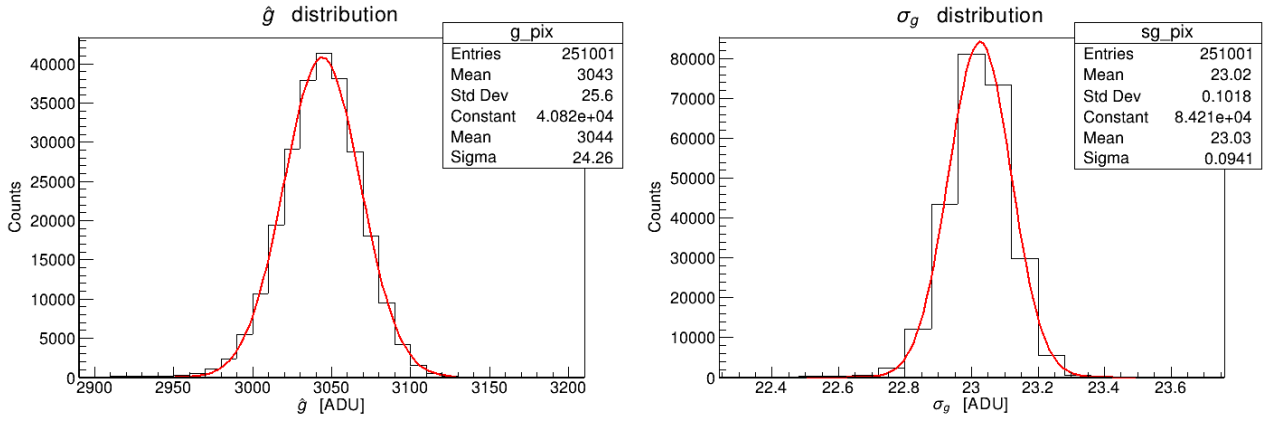


Figura 4.3: Istogrammi di \hat{g} e $\sigma_{\hat{g}}$ relativi all'illuminazione numero 1.

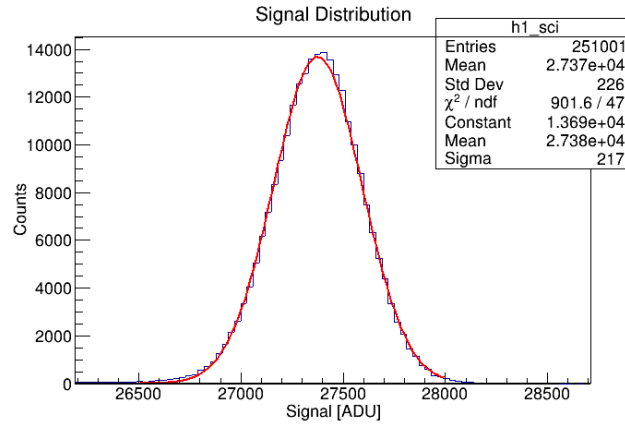


Figura 4.4: Istogramma del segnale integrato del file relativo all'illuminazione 1.

4.3 Errori

Nella sezione precedente abbiamo visto che l'errore su \hat{g} che otteniamo nel caso che abbiamo denominato "no-likelihood" è molto maggiore di quello che ricaviamo negli altri due casi. Inoltre, i risultati della simulazione illustrati in 3.2.2 mostrano come l'errore computato prendendo in considerazione tutte le ipotesi che assumiamo nella formulazione del modello per il calcolo di \hat{g} sia effettivamente quello in miglior accordo con $\text{rms}[\hat{g}]$. Confrontando quindi $\sigma_{\hat{g}}$, $\sigma_{\hat{g}_0}$ e $\text{rms}[\hat{g}]$ calcolando i rapporti $\frac{\text{rms}[\hat{g}]}{\sigma_{\hat{g}}}$ e $\frac{\text{rms}[\hat{g}]}{\sigma_{\hat{g}_0}}$ e il bias tra $\sigma_{\hat{g}}$ e $\sigma_{\hat{g}_0}$, definito come $\frac{\sigma_{\hat{g}_0} - \sigma_{\hat{g}}}{\sigma_{\hat{g}_0}}$. Le simulazioni svolte prevedono che, per flussi superiori ai 5 e⁻/s, il rapporto si stabilizzi a circa 0,9 nel caso $\alpha = 0$ e a 1 nel caso $\alpha \neq 0$, con un possibile bias da questi valori non superiore al 10%. I risultati ottenuti, rappresentati in Figura 4.5, sono in accordo con queste previsioni. Inoltre, esattamente come previsto dalla simulazione, otteniamo un bias del 10% tra gli errori computati nel caso $\alpha = 0$ e $\alpha \neq 0$.

Illuminazione	$\sigma_{\hat{g}}$ [ADU]	$\sigma_{\hat{g}_0}$ [ADU]	$\text{rms}[\hat{g}]$ [ADU]	$\text{rms}[\hat{g}]/\sigma_{\hat{g}}$	$\text{rms}[\hat{g}]/\sigma_{\hat{g}_0}$	$(\sigma_{\hat{g}_0} - \sigma_{\hat{g}})/\sigma_{\hat{g}_0}$
1	23,03	25,39	24,29	1,055	0,946	0,104
2	20,39	22,75	20,42	1,001	0,897	0,104
3	21,15	23,61	22,20	1,050	0,940	0,104

Tabella 4.3: Verifiche dei risultati della simulazione illustrata in 3.2.2.

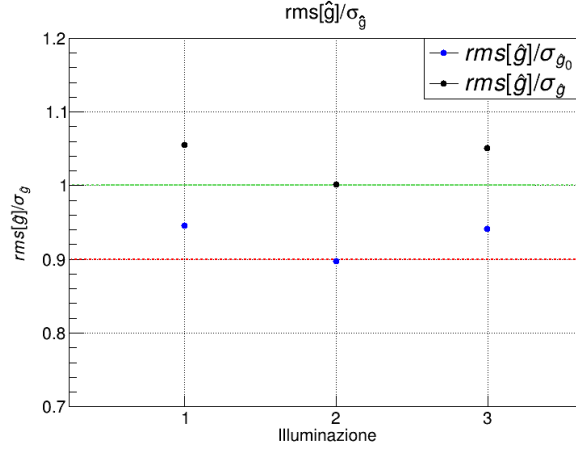


Figura 4.5: Rapporti $\text{rms}[g]/\sigma_{\hat{g}}$ (colore nero) e $\text{rms}[g]/\sigma_{g_0}$ (colore blu). Le linee tratteggiate in verde e in rosso rappresentano i valori previsti dalla simulazione rispettivamente per $\text{rms}[g]/\sigma_{\hat{g}}$ e $\text{rms}[g]/\sigma_{g_0}$.

4.4 Fattori di Qualità

In ultima analisi calcoliamo i fattori di qualità pixel per pixel. Secondo la simulazione illustrata in 3.2.3 i fattori di qualità si distribuiscono come una distribuzione χ^2 a $n_g - 2$ gradi di libertà. Avendo a che fare con esposizioni del tipo MACC(10,16,11) ci aspettiamo, quindi, che la distribuzione dei QF assuma il suo massimo per valori circa pari a 8.

In Figura 4.6 rappresentiamo la distribuzione dei fattori di qualità per l'illuminazione numero 1; tale distribuzione segue il χ^2 , ma presenta il suo massimo per un valore del QF circa pari a 4. Per le alte due illuminazioni, invece, il massimo assume valori vicini a 3, come riportato in Tabella 4.4.

Osserviamo, inoltre, che, come nella simulazione, la distribuzione dei QF nel caso $\alpha = 0$ risulta essere centrata su valori leggermente più bassi rispetto al caso $\alpha \neq 0$.

Illuminazione	QF _{$\alpha \neq 0$}	QF _{$\alpha = 0$}
1	4,11	3,25
2	2,92	2,30
3	3,17	2,58

Tabella 4.4: Massimo delle distribuzioni dei fattori di qualità ottenute nei due diversi casi $\alpha \neq 0$ e $\alpha = 0$.

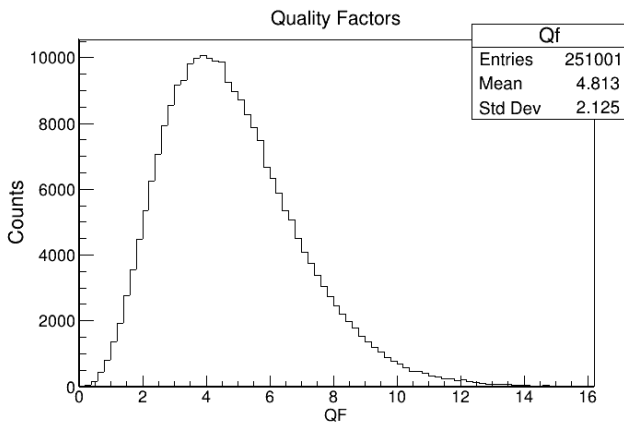


Figura 4.6: Distribuzione dei fattori di qualità per i dati relativi all'illuminazione numero 1.

Inoltre, notando che l'illuminazione 1, quella più intensa tra le tre prese in esame, presenta valori di QF maggiori delle altre due e che la numero 2, la meno intensa risulta avere QF minore tra tutti, siamo portati a supporre che la distribuzione dei QF potrebbe dipendere non solo dal numero di gruppi ma anche dal flusso incidente o dalla modalità di acquisizione.

Al momento, tuttavia, non disponiamo di alcun risultato che giustifichi la sottostima del fattore di qualità che abbiamo riscontrato.

Capitolo 5

Conclusioni

In questa tesi abbiamo presentato un nuovo estimatore per il segnale raccolto dai rivelatori H2RG e per la sua varianza adattato alla modalità di acquisizione dei detector di NISP, e confrontato i risultati delle simulazioni numeriche con i risultati ottenuti applicando tali algoritmi ai dati raccolti da NISP durante le campagne di test dello strumento.

Dall'analisi di alcune immagini acquisite durante una campagna di test per le quali sono disponibili i dati dei singoli gruppi abbiamo visto che i valori di \hat{g} , ossia del segnale accumulato tra due gruppi consecutivi, che abbiamo ottenuto sono, in tutti i casi, in ottimo accordo con la stima di \hat{g} relativa all'immagine *science* e coerente con quanto descritto nell'articolo [8].

Gli unici dati attualmente disponibili sono acquisiti con un flusso troppo intenso per apprezzare alcune caratteristiche salienti dell'algoritmo. Dal confronto delle stime ottenute per le tre diverse espressioni di \hat{g} che abbiamo preso in considerazione siamo in grado di concludere che la stima di \hat{g} che otteniamo con (3.6) è quella che fornisce l'errore minore tra le tre ed è, inoltre, quella il cui errore è in maggiore accordo con la larghezza della distribuzione di \hat{g} ($\text{rms}[\hat{g}]$).

I risultati ottenuti per la distribuzione del quality factor non sono, invece, in accordo con quanto previsto dalla simulazione. È in corso un'indagine per capire quali siano le ragioni di tale sottostima.

Bibliografia

- [1] Euclid Study Team, "*Euclid, Mapping the Geometry of the Dark Universe, Definition Study Report*", European Space Agency, Jul 2011
- [2] G. D. Racca *et al.*, "The Euclid Mission Design", *SPIE Proceedings* Vol. 9904, Jul 2016
- [3] M. Cropper *et al.*, VIS: the visible imager for Euclid, *Proc. SPIE 9904*, 2016
- [4] T. Maciaszek *et al.*, Euclid Near Infrared Spectrometer and Photometer instrument concept and first test results obtained for different breadboards models at the end of phase C, *Proc. SPIE*, 2016
- [5] F. Fornari, "*Software Development and Detector Characterization of the EUCLID Near-Infrared Spectro-Photometer*", tesi di dottorato, Alma Mater Studiorum Università di Bologna, XXX Ciclo, Apr 2018
- [6] C. Bonoli, F. Bortoletto, A. Balestra, M. D'Alessandro, E. Medinaceli, R. Farinelli, "*NI-DPU ASW Data Processing Definition and Justification File*", Euclid Consortium, Dec 2019
- [7] F. Laudisio, "*Verification and integration of the management and control software for the Near Infrared Spectrometer Photometer of the Euclid space mission*", tesi di dottorato, Università degli Studi di Padova, XXXI Ciclo, Nov 2018
- [8] B. Kubik *et al.*, "A New Signal Estimator from the NIR Detectors of the Euclid Mission", *Publications of the Astronomical Society of the Pacific*, Oct 2016
- [9] "The Fits Support Office", <https://fits.gsfc.nasa.gov/>, Dec 2017, ultima consultazione 03/09/2020